# Analyzing Temporal Characteristics of Check-in data

Sushma Bannur
Microsoft Corp.
sushmab@microsoft.com

Omar Alonso
Microsoft Corp.
omalonso@microsoft.com

## ABSTRACT

There is a surge in the use of location activity in social media, in particular to broadcast the change of physical whereabouts. We are interested in analyzing the temporal characteristics of check-ins data from the user's perspective and also at the aggregate level for detecting patterns. In this paper we conduct a large study using check-in data from Facebook to analyze different temporal characteristics in four venue categories (restaurants, movies, shopping, and get-away). We present the results of such study and outline application areas where the conjunction of location and temporal-aware data can help in new search scenarios.

## Categories and Subject Descriptors

H.2.8 [**Database Applications**]: Spatial Databases and GIS; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

## Keywords

Social media, Location Based Social Networks, Facebook

## 1. INTRODUCTION

Social networking sites like Facebook and Twitter are extremely popular as mediums to produce and consume real-time information. Besides following friends and producing status updates, users can also provide location information. This location information can be about the user's default living location or the user's geographic location at a given moment. Most social based services like Facebook, Twitter, and Foursquare provide the ability to define location and to check-in at a particular venue. When a user checks-in somewhere, usually a venue, on a service, friends can know exactly where the user is.

Check-in data can be useful for alerting friends that the user is at the neighborhood and to recommend similar venues based on previous activity, friend's check-in patterns, and global popularity.


*WWW'14 Companion,* April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04
http://dx.doi.org/10.1145/2567948.2579041.

Our interest on these type of data is different: we would like to understand the temporal activity by venue to detect patterns and other characteristics that can be used for building specific applications. Humans are mobile by nature so one can think of scenarios such as local search and mobile location-aware exploratory search where location information can enhance the search user experience. User needs that involve task completion that require spatial and temporal proximity is another scenario that should benefit from the availability of check-in data.

We perform a study using a full year worth of check-in data from Facebook and look at temporal characteristics at the date and seasons level.

In this paper we make the following contributions:

- A large scale analysis of Facebook check-ins in four specific categories.
- Identification of season's patterns per venue.
- A breakdown of check-in activity for the United States and for the states of California and Maine in particular.

This paper is organized as follows. We present a high level overview of related work in the next section. In Section 3, we describe the data set characteristics that we use in our study. In Section 4, we present the analysis and findings. Finally, we present our conclusions and future work.

## 2. RELATED WORK

Exploiting location data from social networks is an active research topic. Most of the work has been done around recommendation, prediction, and personalization of venues based on check-in data. Examples of such work are personalized point-of-interest recommendation (Liu *et al.* [3], Yuan *et al.* [7]) and location recommendation by Gao *et al.* [2].

A couple of studies on Foursquare patterns are the work by Cheng *et al.* [1] and Noulas *et al.* [4], where they analyze the spatio-temporal characteristics of check-ins. Another way of looking at location data is from a city's perspective as investigated by Zhang *et al.* [8] and Silva [6]. A common approach on these studies is the focus on large urban cities. In contrast, we look at check-in data at the country level to understand user behavior.

Our work is closer to the research by Preotiuc-Pietro and Cohn that mine user check-in behavior along with temporal patterns from Foursquare [5]. In our work we use a different data set and make more emphasis on the temporal nature of the user data by going beyond daily activity patterns.

## 3. DATA SET CHARACTERISTICS

In this study we consider all the check-ins in the United States by a sample of users on Facebook in the one year time period from January $1^{st}$ 2013 to December $31^{st}$ 2013. The total number of check-ins in the sampled dataset is about 299.5 Million. Every check-in is associated with a venue name, category of the venue and the timestamp corresponding to when the check-in occurred.

In our data analysis we are primarily interested in check-ins related to restaurants, movies, shopping, and get-away venue categories. We filter the check-ins to group the venues into one of these four categories and discard check-ins that do not belong to any of these categories. Every user check-in on Facebook is associated with a Facebook page and a page is typically associated with a category which is based on a set of category names from a pre-defined list of categories from Facebook. This makes it possible to categorize a user check-in into one of the three categories mentioned above.

Selecting the check-ins corresponding to the movie category is a simple lookup on the Facebook category structure. Specifically, this corresponds to check-ins where the category of the associated Facebook page is *movie theater*.

In the case of restaurants, we consider check-ins associated with venues where the page category is one of: *restaurant*, *restaurant/cafe*, *food & beverage*, or the category name ends with the keyword *restaurant*. Facebook page categories support cuisine based restaurant categorization like *Italian restaurant*, *Indian restaurant*, etc. Thus, the partial match based on the keyword restaurant ensures that we include all the relevant venues.

For shopping, we look at check-ins at venues with category *shopping & retail*, *retail*, and *consumer merchandise* or the category name ends with the keyword store. The partial match using the keyword *store* ensures that we select retailers corresponding to categories like *clothing store*, *outlet store*, etc.

Finally, for the get-away category, we editorially select the relevant category names from the Facebook category structure. The check-in categories corresponding to get-away category in our dataset are: *places & attractions*, *attractions/things to do*, *tours & sightseeing*, *tourist attraction*, *zoo & aquarium*, *cruise excursions*, *winery & vineyard*, *surfing spot*, *theme park*, *national park*, *state park*, *water-park*, *mountain*, *landmark*, *outdoors*, *cruise*, *river*, *beach*, *lake*, and *zoo*.

## 4. DATA ANALYSIS AND RESULTS

In this section we detail our analysis and related findings focusing on the temporal dynamics of check-ins in the four categories that we described in the previous section.

Figure 1 illustrates the distribution of check-ins in the restaurant category across different days and times of a week. The total number of restaurant check-ins in the dataset is about 18.4 Million. We observe that the number of restaurant check-ins over the weekend days is much higher than any weekday, with about 45% of the restaurant check-ins just over the weekend. Moreover, the check-in pattern seems to be biased towards check-ins later in the day compared to mornings.

Figure 2 describes the number of check-ins in the restaurant category across different months in the year. The number of check-ins begin to increase in October, reaching its

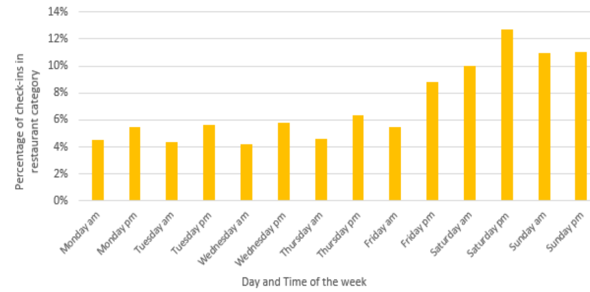


**Figure 1: Distribution of check-ins in the restaurant category across different days & times of a week, based on check-ins in 2013.**

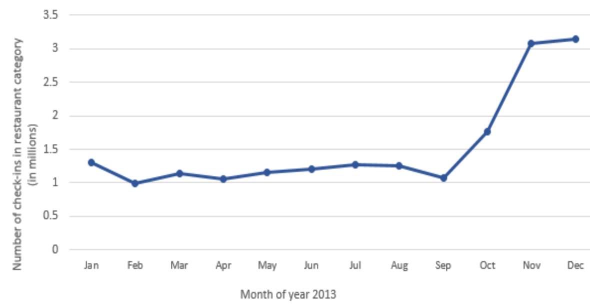


**Figure 2: Check-ins in the restaurant category across different months in 2013.**

peak in December. Besides the rise around the end of the year, the check-in pattern is almost steady around the year with a very slight bump during the summer months.

We perform an identical analysis on check-ins in movie category. The dataset comprises of about 1.2 Million movie check-ins. The number of movie check-ins is far lower in comparison to the restaurant category. Figure 3 outlines the distribution of check-ins in movie category across different days and times of a week.

Similar to the restaurant category, movie check-ins tend to rise around Friday and peak over the weekend, with about half of the movie check-ins occurring during the weekend. This effect seems to be more prominent for movie check-ins as compared to restaurant check-ins.

Figure 4 displays the growth in movie check-ins across different months of the year. Similar to the restaurant cat-

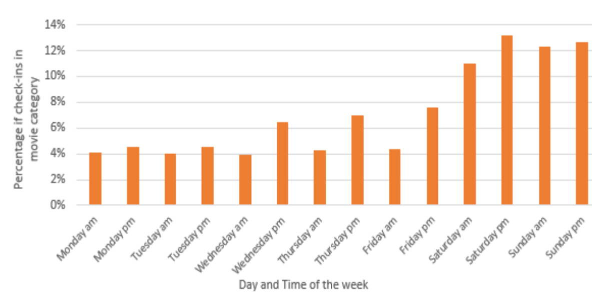


**Figure 3: Check-ins in the movie category across different months in 2013.**

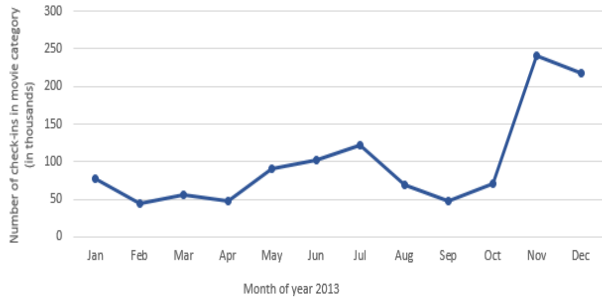


**Figure 4: Check-ins in the movie category across different months in 2013.**

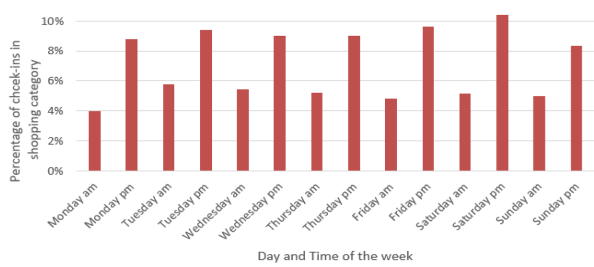


**Figure 5: Distribution of check-ins in the shopping category across different days & times of a week, based on check-ins in 2013.**

egory, the check-ins peak during the holiday season. However, unlike the restaurant category, the check-ins here show a drop in December and a noticeable rise during the summer months. Movie attendance is a leisure activity so this behavior is somewhat expected.

Figure 5 shows the distribution of check-ins in the shopping category across different days and times of a week. The total number of check-ins in the shopping category is around 82.3K. Unlike the check-ins in movies and restaurants categories, the check-ins in the shopping category are almost equally distributed along the week. We see a slight increase on Saturday compared to the rest of the days in the week. However, we observe that people are more likely to shop later in the day compared to the mornings. This effect is similar to restaurants and movies but much stronger here.

The distribution in Figure 5 suggests that most people shop on an ongoing need-basis, unlike a planned restaurant or movie outing over a weekend.

Figure 6 displays the distribution of check-ins in shopping category across different months of the year. The chart clearly indicates that the time period around the holiday season in Christmas and New Year's is when people shop heavily. Apart from this time period, the number of check-ins is about the same along the rest of the year. However, unlike restaurant and movie check-ins, the number of shopping check-ins take a slight dip during the summer months.

Next, we study the geographical and temporal dynamics of check-ins in the get-away category. The dataset comprises of about 19.5 Million get-away check-ins. Figure 7 outlines the number of get-away check-ins across different months in the year. Unlike the restaurant and movie category, get-away check-ins peak during the northern-hemisphere sum-

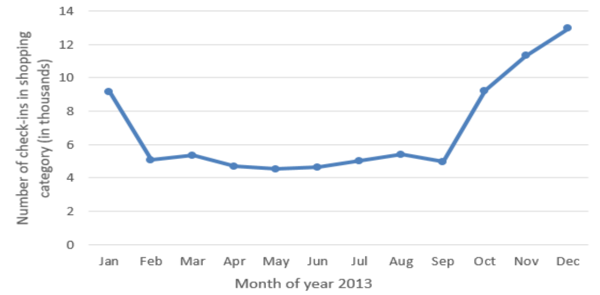


**Figure 6: Check-ins in the shopping category across different months in 2013.**

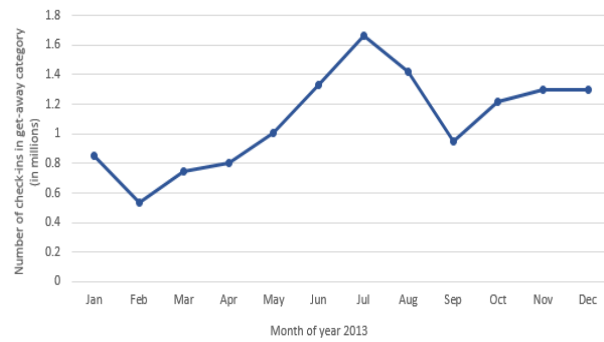


**Figure 7: Check-ins in the get-away category across different months in 2013.**

mer months in June, July, and August followed by the holiday seasons in winters.

Figure 8 plots the get-away check-ins on a daily basis for July, August, and September. Figure 9, plots the same for November and December. The plot in both figures, reveals a rise in the number of check-ins around the weekend and a steep fall right after that. We see in both the plots that the user trails reveal a noticeable increase in check-in activity around the holidays compared to the rest of the days. The peaks marked in red circles in each of these plots reveal all the United States government holidays in the associated time period. Specifically, Figure 8 shows a rise around Independence Day and Labor Day weekends. Check-in plot in Figure 9 indicates an increased activity around Thanksgiving, Veterans Day, and Christmas weekends.

We evaluate the get-away check-ins across the four seasons in the year. Table 1 presents the time period corresponding to each season in the year 2013. In this analysis we compare the variation in check-ins in a given region across the four seasons and also compare the seasonal variation of check-ins

| | |
|---|---|
| Winter | 01/01/2013 to 02/18/2013; 12/01/2013 to 12/31/2013 |
| Spring | 02/19/2013 to 05/25/2013 |
| Summer | 05/26/2013 to 09/05/2013 |
| Fall | 09/06/2013 to 11/30/2013 |

**Table 1: Time period corresponding to the four seasons in the year.**

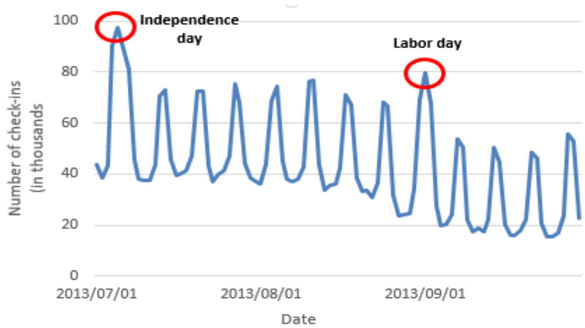


**Figure 8: Daily check-ins in the get-away category in the months of July, August, and September.**

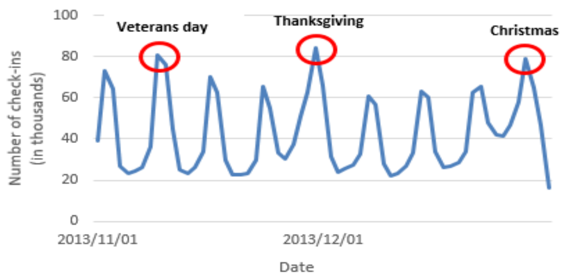


**Figure 9: Daily check-ins in the get-away category in the months of November and December.**

across different regions. One of the challenges in doing such an analysis is that the number of check-ins in a given region is heavily correlated with the density of that region. To address this challenge we define a metric called the *polarity* of a region for a specific season. The polarity of the region for a season is a normalization of the absolute number of check-ins. Polarity is computed as the ratio between the number of get-away check-ins in a given region in a season and the total number of check-ins in that region in the entire year. Specifically, polarity $P$ of a region $r$ for a specific season $s$ is given by:

$$P(r, s) = N(r, s)/N(r)$$

where $N(r, s)$ is the number of check-ins in region $r$ in season $s$ and $N(r)$ is the total number of check-ins in region $r$ across the entire year.

Figure 10 displays the polarity of different states in the United States across the four seasons. Map corresponding to each season has a different coloring scale, where the lightest color corresponds to the region with lowest polarity magnitude in that season and the inverse is true for the darkest shade. As bounds of the polarity values vary across each season, the comparison of a specific region across the four seasons based on Figure 10 is not meaningful.

This figure compares the *variation* in the polarity across different regions in a specific season. For instance, it can be seen that in the winter and spring seasons, warm regions along the west coast and regions like Florida and Hawaii have a much higher polarity compared to the regions in the east coast. Colorado and Vermont are popular ski destinations in the winter, which explains why these cold regions have a much higher winter polarity compared to the other regions in the east coast. On the other hand, cold regions such as Alaska, Maine and Montana show up as regions with high polarity for summer. Nebraska shows up as a region with the highest polarity for fall. Nebraska is located within the country's "tornado alley" and experiences violent thunderstorms and tornadoes throughout the summer and spring seasons and blizzards and ice storms during the winter season. It is fascinating to see how the user-trails capture this phenomena.

Figure 11 illustrates the seasonal polarity of different regions, as defined above, from a contrastive perspective. The color scheme across all four seasons corresponds to the same polarity scale. Therefore, this allows comparison of regions within and across seasons. The regions that demonstrate significant variance in color shade across the four seasons correspond to the ones with strong seasonal preference, whereas the ones with not much change in the color across seasons are the ones with about the same polarity towards each season. For instance, we see that Alaska and Maine show a very strong polarity towards summer, slightly decreased likelihood of being visited in the fall and almost negligible likelihood of being visited in winter and spring. On the contrary, polarity for regions like Hawaii and California demonstrates that these regions are almost equally popular all around the year.

It follows from the above analysis, that the user-trails reveal the seasonal preference for a particular geographic region. To understand this further, we zoom into the venues within states to investigate if the popularity of venues in a given region varies across the four seasons. For this comparison, we limit our analysis to two states, Maine from the east coast and California from the west coast. We select top 10 most popular get-away venues in each season from these two states, popularity defined by the number of check-ins.

We observe that the popular venues for a given region vary across the four seasons, with some venues being popular only in specific seasons. As certain popular venues are missing from the top 10 list in specific seasons, reporting the polarity of a region is not meaningful. Therefore, for this analysis, we look at the absolute number of check-ins for a venue and do not measure the polarity of a venue for a season.

Table 2 shows the top 10 most popular venues, and the associated number of check-ins in each season. The top three most popular venues, Disneyland, Six Flags, and Legoland seem to be the same across the four seasons. However, we see that amongst the top 10 venues, Yosemite National Park (Yosemite NP) is least popular get-away venue in fall and its popularity gradually increases towards winter, spring and summer respectively, making it the fourth popular get-away venue in summer. Another example is the Queen Mary's Dark Harbor (located in Long Beach), which shows up as the fourth most popular venue in fall but does not make it to the top 10 venues for any other season. Queen Mary's Dark Harbor is famous for Halloween celebrations, thus making this venue an appropriate choice for fall season. It is interesting to see that even in a region like California which has about the same polarity for all seasons, the venues exhibit seasonal preferences.

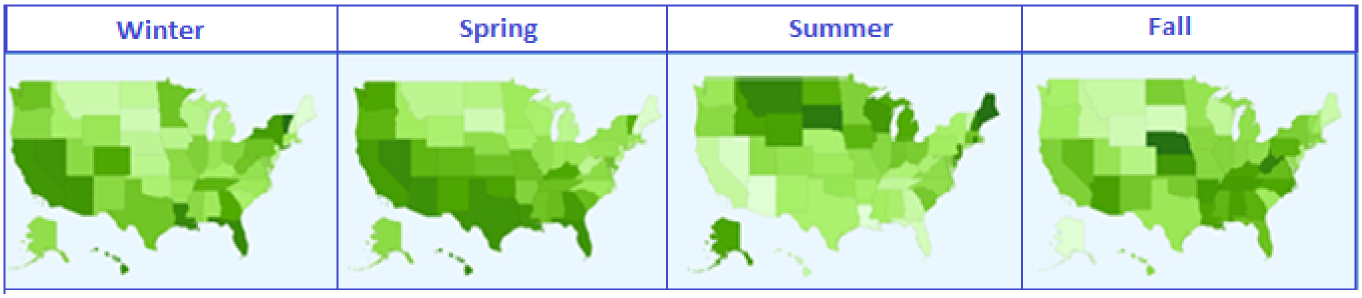


**Figure 10: Seasonal polarity of check-ins in different states in the United States, with the lower and upper bound for coloring scale for each season is defined by the lowest and highest polarity across all regions in a particular season. For example, warm regions along the west coast and regions like Florida and Hawaii have a much higher polarity in winter and spring seasons compared to the regions in the east coast. However, in summer cold regions such as Alaska, Maine, and Montana show up as regions with high polarity.**

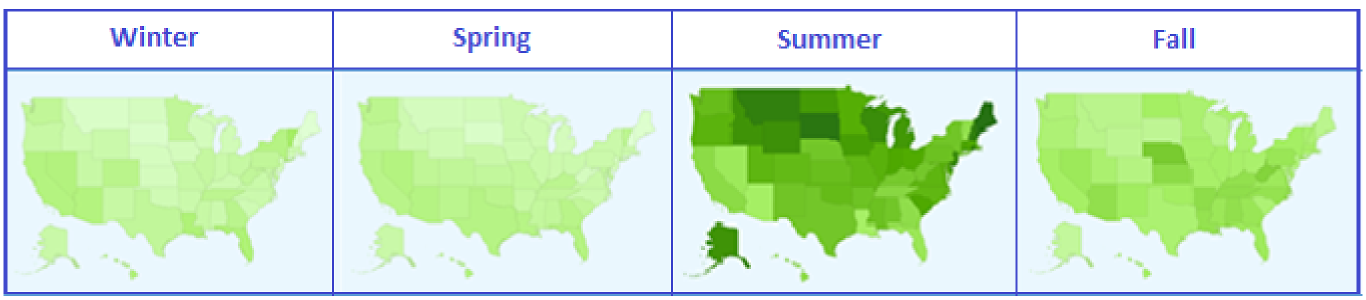


**Figure 11: Seasonal polarity of check-ins in different states in the United States, with the lower and upper bound for coloring scale being same across all seasons; defined by the lowest and highest polarity across all regions considering all seasons. For example, Alaska and Maine show a very strong polarity towards summer and almost nil polarity towards winter and spring. On the contrary, polarity for regions like Hawaii and California remains neutral for all seasons.**

| Winter | | Spring | | Summer | | Fall | |
|---|---|---|---|---|---|---|---|
| Disneyland | 76660 | Disneyland | 67254 | Disneyland | 58911 | Disneyland | 57179 |
| Six Flags | 3109 | Six Flags | 9190 | Six Flags | 13688 | Six Flags | 4932 |
| Legoland | 3070 | Legoland | 2912 | Legoland | 5281 | Legoland | 3051 |
| LA Union Station | 2616 | SD Zoo Safari Park | 2435 | Yosemite NP | 4070 | Queen Mary's Dark Harbor | 2740 |
| SD Zoo Safari Park | 2375 | Yosemite NP | 2288 | La Jolla Cove | 3296 | LA Union Station | 2670 |
| Old Town SD | 2136 | Old Town SD | 2130 | SD Zoo Safari Park | 2373 | Old Town SD | 2616 |
| Venice Beach | 1846 | LA Union Station | 1965 | LA Union Station | 2296 | SD Zoo Safari Park | 2234 |
| Yosemite NP | 1673 | La Jolla Cove | 1828 | SF Zoo | 2068 | Venice Beach | 2112 |
| La Jolla Cove | 1525 | Venice Beach | 1826 | Aquatica SD | 1918 | La Jolla Cove | 1932 |
| SF Zoo | 1409 | SF Zoo | 1634 | Old Town SD | 1872 | Yosemite NP | 1687 |

**Table 2: Top 10 venues (with corresponding check-in counts) in the get-away category in the state of California across different seasons of the year. The first three venues remain popular all year long. Examples of favorites by seasons are Yosemite National Park in the summer and the Queen Mary's Dark Harbor, an annual Halloween tradition, in fall.**

| Winter | | Spring | | Summer | | Fall | |
|---|---|---|---|---|---|---|---|
| Acadia National Park | 194 | Acadia National Park | 292 | Acadia National Park | 2046 | Acadia National Park | 895 |
| Merrill Auditorium | 97 | Short Sands | 73 | Short Sands | 485 | Portland Head Light | 209 |
| Norton Lights Wells | 44 | Portland Head Light | 57 | York Beach | 454 | York Beach | 154 |
| Collins Center for the Arts | 42 | Maine State House | 42 | Portland Head Light | 326 | Merrill Auditorium | 113 |
| Short Sands | 38 | York Beach | 36 | Drakes Island Beach | 212 | Pemaquid Point Light | 76 |
| USA-Canada Border | 29 | Drakes Island Beach | 32 | Goose Rocks Beach | 186 | Ricker Hill Orchards | 74 |
| Higgins Beach | 28 | Baxter State Park | 30 | Pemaquid Point Light | 162 | Baxter State Park | 70 |
| York Beach | 28 | Pemaquid Point Light | 25 | York Wild Animal Kingdom | 149 | USA-Canada Border | 64 |
| Baxter State Park | 28 | Higgins Beach | 24 | Higgins Beach | 147 | Short Sands | 59 |
| Ogunquit | 25 | Goochs Beach | 24 | Ogunquit Beach | 134 | Goose Rocks Beach | 46 |

**Table 3: Top 10 venues (with corresponding check-in counts) in the get-away category in the state of Maine across different seasons of the year. The Acadia National Park is a popular destination all year along peaking around summer. Examples of favorites by seasons are Merrill Auditorium in winter and Short Sands in spring and summer.**

Similar to Table 2, Table 3 lists the top 10 venues for Maine across the four seasons. As demonstrated in the polarity maps, we see that the number of get-away check-ins in summer are significantly larger than the other seasons. Acadia National Park remains the most popular venue across all four seasons. However, popularity of other venues changes significantly across four seasons. For example, Merrill Auditorium is popular only in winter and fall. Another example is, York Wild Animal Kingdom which surfaces in the list of top venues only for summer. The official website for York Wild Animal Kingdom states that this venue is fully functional only in the summer season, which conforms to the information derived from the user check-ins.

# 5. CONCLUSIONS AND FUTURE WORK

We presented a large study of Facebook check-in data with the goal of detecting temporal patterns and dynamics. For popular venues like restaurants and movies, the temporal activity is somewhat expected: mostly evenings and high on the weekends. For shopping, the temporal activity reveals the holiday season as the favorite time for shopping. For other categories like get-away, the temporal characteristics are more prominent, showing very interesting patterns of human mobility with respect to seasons in the United States. We observed the granularity of the check-ins by examining the states of California and Maine in more detail.

From our study we can see that people would check-in on venues of interests (e.g., Yosemite NP) or when performing a social activity like going to movies, which provides valuable insights about their location dynamics. This behavior is very interesting because it can be similar to a query, which in our case is the venue location.

As the Internet keeps evolving, we expect to see new location services and changes in use and adoption. In our work we use data from a popular social network service which is a good sample of the overall population.

Future work includes expanding the set of categories and countries and replicating the methodology with a different location data set with the goal of generalizing the findings. Another line of work is measure the proximity of venues in those categories and better ways to visualize the data.

Check-in data offers lots of opportunities for designing new search experiences that help users with their information needs in novel ways. We believe that understanding users from their check-in logs can be useful for building new applications that are temporally and location aware.

# 6. REFERENCES


[1] Z. Cheng, J. Caverlee, K. Lee, and D. Sui. Exploring Millions of Footprints in Location Sharing Services. In *Proc. of ICWSM* 2011.

[2] H. Gao, J. Tnag, X. Hu, and H. Liu. Exploring Temporal Effects for Location Recommendation on Location-Based Social Networks. In *Proc. of RecSys* 2013.

[3] X. Liu, Y. Liu, K. Aberer, and C. Miao. Personalized Point-of-Interest Recommendation by Mining User's Preference Transition. In *Proc. of CIKM* 2013.

[4] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil. An Empirical Study of Geographic User Activity Patterns in Foursquare. In *Proc. of ICWSM* 2011.

[5] D. Preotiuc-Pietro and T. Cohn. Mining User Behaviours: A Study of Check-in Patterns in Location Based Social Networks. In *Proc. of WebSci* 2013.

[6] T. Silva, P. Vaz de Melo, J. Almeida, J. Salles, and A. Loureiro. Visualizing the invisible image of cities. In *Proc. of GreenCom* 2012.

[7] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. Magnenat-Thalmann. Time-aware Point-of-interest Recommendation. In *Proc. of SIGIR* 2013.

[8] K. Zhang, Q. Jin, K. Pelechrinis, and T. Lappas. On the Importance of Temporal Dynamics in Modeling Urban Activity. In *Proc. of UrbComp* 2013.